

Handbook of Positive Psychology Assessment

**Willibald Ruch
Arnold B. Bakker
Louis Tay
Fabian Gander**
(Editors)

Psychological Assessment –
Science and Practice

 **hogrefe**

Handbook of Positive Psychology Assessment

About the Editors

Willibald Ruch, PhD, is Professor for Personality and Assessment in the Department of Psychology at the University of Zurich, Switzerland. He is a fellow of the International Positive Psychology Association, a senior scientist at the VIA Institute of Character in Cincinnati, OH, USA, and past president of the Swiss Association for Positive Psychology and the International Society for Humor Studies. His research interests concern issues of defining and measuring personality and character, in particular character strengths, virtues, humor, cheerfulness, and positive emotions.

Arnold B. Bakker, PhD, is Professor of Work and Organizational Psychology at Erasmus University Rotterdam, The Netherlands, as well as a (distinguished) visiting professor at the University of Johannesburg and North-West University, South Africa, the University of Zagreb, Croatia, and the University of Bergen, Norway. He is a fellow of the Association for Psychological Science, the International Association of Applied Psychology, and the European Academy of Occupational Health Psychology as well as former president of the European Association of Work and Organizational Psychology. His research interests cover positive organizational phenomena such as job demands-resources theory, work engagement, flow, playful work design, and job crafting.

Louis Tay, PhD, is William C. Byham Associate Professor of Industrial-Organizational Psychology at Purdue University, West Lafayette, IN, USA. He obtained his PhD in industrial-organizational psychology at the University of Illinois at Urbana-Champaign. His research interests cover methodology (i.e., measurement, continuum specification, latent class modeling, Big Data / data science) and well-being (i.e., societal well-being, wellness programs, work-leisure [e.g., arts/humanities activities] interface).

Fabian Gander, PhD, is a postdoc in the Department of Psychology at the University of Basel, Switzerland. He received his PhD from the University of Zurich and he serves as co-editor of the *Journal of Happiness Studies*. His research interests are personality traits and well-being, with a special focus on changes in personality and well-being and their assessment.

Psychological Assessment – Science and Practice

Each volume in the series *Psychological Assessment – Science and Practice* presents the state-of-the-art of assessment in a particular domain of psychology, with regard to theory, research, and practical applications. Editors and contributors are leading authorities in their respective fields. Each volume discusses, in a reader-friendly manner, critical issues and developments in assessment, as well as well-known and novel assessment tools. The series is an ideal educational resource for researchers, teachers, and students of assessment, as well as practitioners.

Psychological Assessment – Science and Practice is edited with the support of the European Association of Psychological Assessment (EAPA).

Editor-in-Chief: Tuulia M. Ortner, Austria

Editorial Board: Itziar Alonso-Arbiol, Spain; Samuel Greiff, Luxembourg; Willibald Ruch, Switzerland; Karl Schweizer, Germany

Psychological Assessment – Science and Practice, Vol. 5

Handbook of Positive Psychology Assessment

Edited by

**Willibald Ruch, Arnold B. Bakker, Louis Tay,
and Fabian Gander**

Library of Congress Cataloging in Publication information for the print version of this book is available via the Library of Congress Marc Database under the LC Control Number 2022942205

Library and Archives Canada Cataloguing in Publication

Title: Handbook of positive psychology assessment / edited by Willibald Ruch, Arnold B. Bakker, Louis Tay, and Fabian Gander.

Names: Ruch, W. (Willibald), editor. | Bakker, Arnold B., editor. | Tay, Louis (Psychologist), editor. | Gander, Fabian, editor.

Series: Psychological assessment--science and practice ; v. 5.

Description: Series statement: Psychological assessment--science and practice ; v. 5 | Includes bibliographical references.

Identifiers: Canadiana (print) 20220398461 | Canadiana (ebook) 20220398577 | ISBN 9780889376199 (softcover) | ISBN 9781616766191 (PDF) | ISBN 9781613346198 (EPUB)

Subjects: LCSH: Positive psychology. | LCSH: Psychological tests—Handbooks, manuals, etc.

Classification: LCC BF204.6 .H36 2022 | DDC 150.19/88—dc23

© 2023 by Hogrefe Publishing

<http://www.hogrefe.com>

The authors and publisher have made every effort to ensure that the information contained in this text is in accord with the current state of scientific knowledge, recommendations, and practice at the time of publication. In spite of this diligence, errors cannot be completely excluded. Also, due to changing regulations and continuing research, information may become outdated at any point. The authors and publisher disclaim any responsibility for any consequences which may follow from the use of information presented in this book.

Registered trademarks are not noted specifically as such in this publication. The use of descriptive names, registered names, and trademarks does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

PUBLISHING OFFICES

USA: Hogrefe Publishing Corporation, 44 Merrimac St., Suite 207, Newburyport, MA 01950
Phone (978) 255 3700; E-mail customersupport@hogrefe.com

EUROPE: Hogrefe Publishing GmbH, Merkelstr. 3, 37085 Göttingen, Germany
Phone +49 551 99950-0, Fax +49 551 99950-111; E-mail publishing@hogrefe.com

SALES & DISTRIBUTION

USA: Hogrefe Publishing, Customer Services Department,
30 Amberwood Parkway, Ashland, OH 44805
Phone (800) 228-3749, Fax (419) 281-6883; E-mail customersupport@hogrefe.com

UK: Hogrefe Publishing, c/o Marston Book Services Ltd., 160 Eastern Ave., Milton Park,
Abingdon, OX14 4SB

Phone +44 1235 465577, Fax +44 1235 465556; E-mail direct.orders@marston.co.uk

EUROPE: Hogrefe Publishing, Merkelstr. 3, 37085 Göttingen, Germany
Phone +49 551 99950-0, Fax +49 551 99950-111; E-mail publishing@hogrefe.com

OTHER OFFICES

CANADA: Hogrefe Publishing, 82 Laird Drive, East York, Ontario, M4G 3V1

SWITZERLAND: Hogrefe Publishing, Länggass-Strasse 76, 3012 Bern

Copyright Information

The eBook, including all its individual chapters, is protected under international copyright law. The unauthorized use or distribution of copyrighted or proprietary content is illegal and could subject the purchaser to substantial damages. The user agrees to recognize and uphold the copyright.

License Agreement

The purchaser is granted a single, nontransferable license for the personal use of the eBook and all related files.

Making copies or printouts and storing a backup copy of the eBook on another device is permitted for private, personal use only. This does not apply to any materials explicitly designated as copyable material (e.g., questionnaires and worksheets for use in practice).

Other than as stated in this License Agreement, you may not copy, print, modify, remove, delete, augment, add to, publish, transmit, sell, resell, create derivative works from, or in any way exploit any of the eBook's content, in whole or in part, and you may not aid or permit others to do so. You shall not: (1) rent, assign, timeshare, distribute, or transfer all or part of the eBook or any rights granted by this License Agreement to any other person; (2) duplicate the eBook, except for reasonable backup copies; (3) remove any proprietary or copyright notices, digital watermarks, labels, or other marks from the eBook or its contents; (4) transfer or sublicense title to the eBook to any other party.

These conditions are also applicable to any files accompanying the eBook that are made available for download.

Should the print edition of this book include electronic supplementary material then all this material (e.g., audio, video, pdf files) is also available with the eBook edition.

Format: PDF

ISBN 978-0-88937-619-9 (print) • ISBN 978-1-61676-619-1 (PDF) • ISBN 978-1-61334-619-8 (EPUB)

<https://doi.org/10.1027/00619-000>

Contents

Part I Introduction

Chapter 1	Introduction	3
	<i>Willibald Ruch, Arnold B. Bakker, Louis Tay, and Fabian Gander</i>	

Part II Well-Being

Chapter 2	Assessing Psychological Flourishing: A Review of Theory and Instruments	7
	<i>Fanyi Zhang and Louis Tay</i>	
Chapter 3	Assessing Subjective Well-Being: A Review of Common Measures ...	38
	<i>William Tov, Jun Sheng Keh, Yan Qiang Tan, Qin Ying Tan, and Indra Alam Syah Bin Aziz</i>	
Chapter 4	Measuring Positive Affect, Positive Emotions, and Positivity Resonance	58
	<i>Jieni Zhou and Barbara L. Fredrickson</i>	
Chapter 5	Health-Related Quality of Life	82
	<i>Stefan Höfer</i>	

Part III Positive Traits, States, and Behaviors

Chapter 6	Assessing Character Strengths	101
	<i>Lisa Wagner and Willibald Ruch</i>	
Chapter 7	Assessing Humor and Its Kin	122
	<i>Sonja Heintz and Jennifer Hofmann</i>	
Chapter 8	Assessing Playfulness: Current Challenges and Overview.	145
	<i>René T. Proyer and Kay Brauer</i>	
Chapter 9	Measuring Meaning and Purpose in Life: A Review and Recommendations	162
	<i>Michael F. Steger and Frank Martela</i>	
Chapter 10	Assessing Flow: State of the Art and Challenges for the Future	175
	<i>Antonella Delle Fave and Marta Bassi</i>	
Chapter 11	Self-Efficacy	207
	<i>Ralf Schwarzer and Aleksandra Luszczynska</i>	
Chapter 12	Assessing the Trait of Appreciation of Beauty	218
	<i>Rhett Diessner</i>	
Chapter 13	Assessing Posttraumatic Growth	234
	<i>Sara Etz Mendonca, A. Demaske, Emma Sims, and Eranda Jayawickreme</i>	

Chapter 14	On the Passion Scale: Theory, Research, and Psychometric Properties	248
	<i>Robert J. Vallerand and Sonia Rahimi</i>	
Chapter 15	Work Engagement: A Critical Assessment of the Concept and Its Measurement	273
	<i>Wilmar B. Schaufeli and Arnold B. Bakker</i>	
Part IV Applied Settings		
Chapter 16	School-Based Approaches for the Universal Assessment of Adolescent Psychosocial Strengths	299
	<i>Jennica Paz, Eui Kyung Kim, Erin Dowdy, Michael J. Furlong, Tameisha Hinton, José A. Piqueras, Tíscar Rodríguez-Jiménez, Juan C. Marzo, and Susan Coats</i>	
Chapter 17	Dyadic Coping as a Concept of Positive Psychology	317
	<i>Mirjam Senn and Guy Bodenmann</i>	
Chapter 18	Health and Clinical Settings	336
	<i>Carmelo Vazquez, Gonzalo Hervas, Cristian Ochoa, Covadonga Chaves, and Pablo Roca</i>	
Chapter 19	Measuring Leisure: Synopsis of Perspectives and Instruments	351
	<i>Lynn A. Barnett and Joel A. Blanco</i>	
Chapter 20	Monitoring Well-Being Changes in Positive Psychology Interventions	368
	<i>Fabian Gander</i>	
Part V New Directions		
Chapter 21	Measuring Primal World Beliefs	385
	<i>Jeremy D. W. Clifton</i>	
Chapter 22	Imagination	402
	<i>Darya L. Zabelina</i>	
Chapter 23	Assessing Self-Transcendent Experience	418
	<i>David Bryce Yaden and Marianna Graziosi</i>	
Chapter 24	Measuring Nostalgia	443
	<i>Tim Wildschut and Constantine Sedikides</i>	
Contributors		459

Part I

Introduction

Chapter 1

Introduction

**Willibald Ruch¹, Arnold B. Bakker², Louis Tay³,
and Fabian Gander⁴**

¹Department of Psychology, University of Zurich, Switzerland

²Center of Excellence for Positive Organizational Psychology,
Erasmus University Rotterdam, The Netherlands

³Department of Psychological Sciences, Purdue University,
West Lafayette, IN, USA

⁴Department of Psychology, University of Basel, Switzerland

Since the advent of positive psychology around the turn of the millennium, research and practice in this area have flourished. Not only has research into existing positive concepts increased but numerous new concepts have also been introduced and new assessment instruments and methods have been developed. For many topics, this has led to a plethora of – often competing – approaches to measurement. Today, researchers and practitioners alike are often faced with the challenging task of finding their way through a maze of alternative approaches when aiming to assess a particular concept. In addition, relatively little research explicitly addresses diagnostic issues, compares instruments, or even offers specific guidelines and recommendations about which measure is particularly suitable for which situation.

This handbook aims to relieve that predicament by providing a state-of-the-art overview of current theories, approaches, issues, and assessment instruments in the field of positive psychology. It is aimed at researchers, instructors, students, and practitioners and serves to guide both researchers and practitioners in selecting appropriate instruments by providing specific recommendations. Thus, the book's overarching goal is to contribute to both theory *and* practice of positive psychological assessment and stimulate further advances in the field by illuminating current gaps in the literature and discussing general issues in the assessment of positive psychological concepts.

Of course, given the breadth of the field and the numerous existing concepts and measurement approaches, this handbook cannot provide an exhaustive overview of the field but rather must be selective. In our selection of topics, we aimed to both cover rather traditional positive psychological concepts and include comparatively new and emerging ones as well. We believe this approach provides readers with the foundational positive psychological concepts while also introducing more novel perspectives.

The chapters are authored by renowned experts in their field. The authors were asked to describe their own work as well as other important contributions to the respective topic. Also, they were invited to not just give a purely neutral and descriptive view of their field

but to include their expert evaluations and opinions on the topic to provide some guidance for the interested reader.

Each chapter begins with an introduction to the theoretical background, which elaborates on the relevance of the topic at hand, followed by an overview of the most relevant assessment instruments in the field, including a discussion of their psychometric properties and a selection of key research findings. Finally, each chapter discusses specific assessment-related challenges regarding the respective topic and provides recommendations for selecting assessment instruments.

The book is divided into four main sections. The first section focuses on well-being. Given the large number of competing theories, models, and assessment instruments on well-being and related concepts (e.g., happiness, flourishing, thriving, positive affect, quality of life), we deemed a current overview of existing approaches to be urgently needed.

The second section of the book covers traits, states, and behaviors. In this section, we had to be the most selective and decided to focus on certain specific topics and cover them in considerable detail: character, humor, playfulness, meaning and purpose, flow, self-efficacy, appreciation of beauty, posttraumatic growth, passion, and work engagement.

The third section of the book focuses on assessment in specific contexts, namely, in school settings, romantic relationships, health and clinical settings, leisure, and positive psychology interventions.

The fourth and final section covers topics that have recently been introduced or have yet to be considered from a positive psychology perspective: primal world beliefs, imagination, self-transcendent experiences, and nostalgia.

Acknowledgments

We thank everyone who contributed to the creation of this book. Foremost, of course, we acknowledge the invaluable contributions of the authors of the individual chapters, who invested their effort and expertise in creating comprehensive overviews of the role of psychological assessment in their respective field. Furthermore, we are very grateful for the contributions of numerous anonymous reviewers who provided critical feedback on the manuscripts and thereby helped to improve the quality of the individual chapters.

Part II

Well-Being

Chapter 2

Assessing Psychological Flourishing

A Review of Theory and Instruments

Fanyi Zhang and Louis Tay

Department of Psychological Sciences, Purdue University,
West Lafayette, IN, USA

In psychology, the concept of flourishing is often mentioned in the same breath as positive psychology (Seligman & Csikszentmihalyi, 2000). The hallmarks of flourishing – positive experiences, positive individual traits, and positive institutions – are aspirational goals and key topics of study in positive psychology. As individuals and societies seek to flourish, they recognize that economic metrics, while important, need to be supplemented by other metrics that directly index human flourishing (Diener & Seligman, 2004). Many of these indices and assessments are directed toward the dimension of positive experiences – which we term psychological flourishing (e.g., Diener, 2000; Su et al., 2014). In this review, we examine the concept of psychological flourishing and the more established major instruments used to assess it.

Human Flourishing and Psychological Flourishing

The concept of human flourishing can be traced back to the Greek concept of *eudaimonia*, which has been translated as happiness, human welfare, and – pertinent to this chapter – human flourishing. Human flourishing points to the highest human good and an objectively desirable life. Indeed, Aristotle noted that human flourishing is “something complete and self-sufficient, since it is the end of the things achievable in action” (Aristotle & Irwin, 1999, p. 8). Yet, what exactly comprises human flourishing? Based on positive psychology (Seligman & Csikszentmihalyi, 2000) and positive health (Seligman, 2008), the term denotes multiple senses: positive individual traits, positive physical health, positive institutions, and positive experiences.

In terms of *positive individual traits*, the Aristotelian conception of human flourishing emphasizes living in accordance with the highest virtue (*arête*; Aristotle & Irwin, 1999). Historically, moral psychology sought to examine this through the efforts of cognitive approaches. For example, in Kohlberg’s theory of moral development (Kohlberg, 1958),

moral reasoning and judgment are the means for evaluating moral growth. More recently, affective approaches have increased in popularity. Discrete emotions such as disgust undergird violations of moral offenses (Rozin et al., 2009) and awe can inspire virtuous action (Keltner & Haidt, 2003; Yaden et al., 2018). Work on character strengths and virtues (Peterson & Seligman, 2004) has resulted in assessments of positive individual attributes and actions (McGrath, 2014; Ng et al., 2018).

The concept of *positive physical health* was recognized by the World Health Organization in 1948 as denoting not merely the absence of illness or infirmity but a complete state of physical wellness, an idea that has since been extended into the field of positive health, which proposes the assessments of healthy functioning (Seligman, 2008). *Positive institutions* include the positive functioning of communities, businesses, and organizations, which means seeking structures and policies that promote fairness and inclusion and enabling collective well-being, for example, in the form of organizational policies that build interpersonal trust (Six & Sorge, 2008).

Our chapter focuses on psychological flourishing, which is the dimension of *positive experiences* in positive psychology. Broadly understood, this comprises both positive subjective experiences (e.g., happiness, flow) and positive interpersonal relationships (e.g., friendships) (see also Park et al., 2016). Psychological flourishing has its roots in the humanistic movement within psychology. The humanists emphasized self-actualization, which refers to the optimal functioning of a person (Maslow, 1956; Rogers, 1961). Indeed, self-actualization, according to the hierarchy of needs, emphasizes the fulfillment of psychological needs (i.e., esteem needs and belongingness needs) beyond physical needs to achieve one's full potential (Maslow, 1943, 1956). Often, psychology has instantiated dimensions of psychological flourishing as both subjective well-being (i.e., positive emotions, low negative emotions, and life satisfaction) (Diener, 1984) and psychological well-being (i.e., self-acceptance, environmental mastery, positive relations, purpose in life, personal growth, and autonomy) (Ryff & Keyes, 1995). Others consider this both hedonic and eudaimonic well-being (Waterman, 1993). Therefore, the concept of psychological flourishing goes beyond positive emotions alone (e.g., Tay et al., 2019), although these are important in their own right. Within positive psychology, psychological flourishing has its incarnation in the PERMA notion of well-being, an acronym for **p**ositive emotions, **e**ngagement, **p**ositive relationships, **m**eaning, and **a**ccomplishment (Seligman, 2011).

We distinguish the concept of human flourishing from psychological flourishing: The former comprises positive actions, attributes, and experiences for entities at different levels of analysis (e.g., individual, group, community, organization, nation) (e.g., Tay et al., 2018), whereas the latter is a subset of human flourishing. The focus lies on the individual experience of positivity, which goes beyond the mere experience of pleasure to encompass positive psychological fulfillment.

Assessing Psychological Flourishing

There are many different means of assessments within and outside of psychology to evaluate specific aspects of psychological flourishing. For example, researchers have developed assessments for meaning in life (Steger et al., 2006), positive affect (Watson et al., 1988), and social support (Zimet et al., 1988). In the assessment of psychological flourish-

ishing, we want to review measures that capture multiple constructs rather than single-ones or that encompass multiple aspects in terms of their content. There are several reasons for this. It is increasingly being recognized that the assessment of human flourishing must be approached in an integrative rather than a piecemeal fashion (VanderWeele, 2017; VanderWeele et al., 2020) to enable communities and societies to better understand the different aspects of human flourishing simultaneously. Methodologically, the use of measures with the same wording style and response scale enables a level of stylistic and rating equivalence (Tay et al., 2021). This is important for researchers when they compare different constructs. For example, past work has shown that measurement of the same constructs on different scale lengths can lead to nonequivalence – even when using linear-stretch methods to place them on a common scale (Batz et al., 2016).

Below, we review psychological flourishing scales that have been validated and well-received in the field, including broad scales that assess multiple dimensions: the Mental Health Continuum (MHC), the Flourishing Scale (FS), the PERMA-Profiler, and the Comprehensive Inventory of Thriving (CIT). Further, we review both Ryff’s Psychological Well-Being Scale (PWBS) and the Personal Well-Being Index (PWI), widely used measures that assess different dimensions of psychological wellness and life domains. At the time of this review, the original works of each scale had been cited over 300 times. We provide a general description of the scales, the key dimensions they assess, and their psychometric properties, including reliability, (factorial) validity, and translation/equivalence across cultures. This chapter draws on studies that focus on examining the psychometric properties of the scales (i.e., reliability, validity, measurement equivalence). We conducted a literature search in ERIC, Google Scholar, PsycARTICLES, PsycINFO, and Social Sciences Full-Text databases using the applicable scale name in combination with the terms “psychometrics” or “validation” as keywords. We also checked the reference lists from the retrieved studies. This initial search yielded 235 articles as well as data presented at professional conferences. We examined and included contributions if they (1) were empirical, (2) aimed at developing or validating the target scale, and (3) provided information on at least one of the psychometric properties mentioned above. The final list of articles utilized in the current review numbered 190 studies; a list of the studies can be found in a supplement at www.wam-lab.com.

Review of Psychological Flourishing Scales

Mental Health Continuum-Short Form (MHC-SF)

Keyes (2005) outlined the two-continua model, where mental health and mental illness were hypothesized to be two related but distinct continua. More specifically, the paper pointed out that the absence of mental illness is not necessarily evidence of high levels of mental health. Based on a review of the previous well-being literature, the state of mental health was operationalized as “a syndrome of a set of symptoms of an individual’s subjective well-being,” with “subjective well-being” referring to the composite of affective states, psychological well-being, and social functioning. The conceptualization and measurement of affective states followed the hedonic approach (e.g., Diener et al., 1999) to capture individuals’ positive feelings and emotions. The latter two followed the eudaimonic approach and focused on two important aspects of optimal functioning. Psy-

chological well-being targeted one's own evaluation of functioning in life, based on Ryff's (1989) characterization of a positive psychological functioning that covered self-acceptance, personal growth, positive relationships with others, environmental mastery, purpose in life, and autonomy. Finally, social functioning focused on how people evaluated their functioning in life in terms of the social standard, based on the rationale in Keyes (1998), which included social contribution, social acceptance, social actualization, social integration, and social coherence.

Originally, researchers used the Mental Health Continuum-Long Form (MHC-LF) to measure these individual aspects, but it was a slightly lengthy scale that consisted of 40 items in total (Keyes, 2002). The Mental Health Continuum-Short Form (MHC-SF) was thus developed to serve as a brief questionnaire to cover the hypothesized three aspects of mental health: emotional (EWB, 3 items), social (SWB, 5 items), and psychological well-being (PWB, 6 items). Respondents are asked to rate the frequency of their feelings on a scale from 1 (*never*) to 6 (*every day*). As a diagnostic tool, the results are categorized into three levels of well-being: flourishing (those who rate 5 or 6 on at least 1 EWB item and at least 6 SWB and PWB items), languishing (those who rate 1 or 2 on at least 1 EWB item and at least 6 SWB and PWB items), and moderate (those who are neither flourishing nor languishing). We found 59 articles in our initial literature search. After excluding those primarily elaborating on theoretical backgrounds, we included 47 in our final review.

Reliability

The initial studies with U.S. adolescents reported acceptable Cronbach's alphas for the subscales (.67 to .84 range; Keyes, 2006). Further studies on Western populations (Canadian and U.S. adults) reported similar data (.77 to .87 range; Gilmour, 2014; Keyes et al., 2012; Orpana et al., 2017).

Hides et al. (2016) reported McDonald's omega coefficients to further validate the subscales. The MHC-SF overall score showed high reliability, with omega coefficients reaching .96 and .90 for classical and hierarchical tests, respectively. The classical omega tests for the subscales ranged from .89 to .91, which appeared to be acceptable. However, the specific omega hierarchical scores for subscales ranged from .03 to .23, none of which met the cutoff of .50 suggested by Reise (2012). These statistics indicated that the variance attributed to the subscales was on the lower end, which was also supported in further studies on more diverse samples. We did not find any temporal reliability information using Western adult samples.

Validity

The MHC-SF was originally validated on a U.S. adolescent sample (Keyes, 2006). PWB correlated positively with self-concept ($r = .54$) and self-determination ($r = .46$); SWB correlated with school integration ($r = .42$), and perceived closeness ($r = .31$); EWB correlated negatively with depression ($r = -.30$). All subscales correlated negligibly with perceived math and reading skills ($r = .13$ to $.22$), showing discriminant validity evidence. Moreover, 90 % of the studies agreed upon a correlated two-factor structure (compared with

one- or orthogonal two-factor structure) between mental health and mental illness ($r = -.34$ to $-.84$ between latent factors), supporting Keyes' overarching two-continua model of human flourishing.

Regarding MHC-SF factorial validity evidence, early researchers tested the proposed three-factor CFA model (Keyes, 2002, 2006; Keyes et al., 2008). In a comparison of one-, two-, and three-factor CFA results, 22 of the 22 factorial validation studies showed that the correlated three-factor CFA yielded the best model fit. The observed interfactor correlations ranged from .53 to .94, 60 % of which were above .70. This raises concerns over the distinctiveness of the three subscales. Because MHC-SF may be viewed as capturing a broad psychological flourishing factor, researchers then hypothesized a bifactor structure (i.e., a broad psychological flourishing factor and three subfactors). 80 % of the studies demonstrated a better model fit of bifactor (with one general factor and three group factors) over three-factor models, while the remaining two studies on the Spanish version showed comparable model fit indices between the two (Echeverria et al., 2017; Peña Contretras et al., 2017). The item-level analysis also supports a broad psychological flourishing factor (e.g., Lamborn et al., 2018; Longo et al., 2020; Rogoza et al., 2018; Schutte & Wissing, 2017; Silverman et al., 2018). The overall item loadings suggest that items loaded more strongly on the general factor rather than their target factor. These results again indicate the presence of a strong general flourishing factor.

Applications in Diverse Samples

The MHC-SF has been tested in 52 countries in over 30 languages. Two studies reported the temporal stability of the MHC-SF. One study with Dutch adults (Lamers et al., 2011) reported stable reliability of the total scale over 9 months, reaching a .65 test-retest correlation. The temporal stability of the subscales ranged from .46 to .53, which also suggested good temporal stability. Further, the test-retest correlations over 3 months ranged from .65 to .70 for the total scale and .45 to .56 across all three subscales. Another study with Italian adults (Petrillo et al., 2015) suggested only moderate temporal reliability over a 1-month interval, with correlations ranging from .27 to .32.

Reliability information for the MHC-SF was reported in 42 studies that examined diverse samples. Cronbach's alpha ranged from .72 to .96 for the overall score, with 94.1 % of the statistics above .80. The subscales exhibited slightly lower but good alphas: EWB had a range from .70 to .92, PWB a range from .66 to .93. The lowest among the three, SWB, ranged from .49 to .88, with 92.5 % of the statistics above .70, which can be considered acceptable (Nunnally & Bernstein, 1994).

Six articles further tested McDonald's omega. The MHC-SF overall score again showed high reliability, with omega coefficients (both classical and hierarchical tests) ranging from .74 to .95. When the results were replicated in Western cultures, while the classical omega tests for the subscales appeared to be acceptable (.55 to .94, with only one exception below .50 cutoff with Setswana-speaking African adults; Schutte & Wissing, 2017), the specific omega hierarchical scores for subscales ranged from .00 to .38, none meeting the cutoff. These statistics again suggested that, while the MHC-SF total score might be treated as a highly reliable measure of positive mental health, subscale score reliability should be considered carefully, as most of the variance was attributable to the gen-

eral factor; the ability of the subscales to reliably measure the specific domains of psychological flourishing beyond the general dimension is relatively low.

The overall MHC-SF positively correlated with mental health ($r = .40$ to $.82$) and subjective happiness ($r = .57$ to $.78$). MHC-SF negatively correlated with psychological distress ($r = -.50$ to $-.70$), depression ($r = -.67$ to $-.38$), anxiety ($r = -.49$ to $-.36$) and stress ($r = -.60$ to $-.41$). The MHC-SF subscales of EWB were positively correlated with life satisfaction ($.49$ to $.65$); PWB was positively correlated with self-esteem ($r = .33$ to $.69$), and SWB was positively correlated with sense of belonging ($r = .40$ to $.50$). The only exception was a study with Indian adolescents, which showed only nonsignificant or weak correlations ($r = -.34$ to $-.10$) with other constructs (Singh et al., 2015).

A total of 19 studies reported measurement invariance over sex, age, countries, ethnicity, education, and geographic location. All 14 studies reached scalar invariance across both sexes, meaning that scores can be compared between males and females. 80 % of the studies reached scalar invariance across age, with one study on Portuguese children reaching only metric invariance (Carvalho et al., 2016). Three studies reported partial scalar invariance across countries. Joshanloo et al. (2013) identified four noninvariant items being items 1 ("How often did you feel happy?"), 4 ("How often did you feel that you had something important to contribute to society?"), 8 ("How often did you feel that the way our society works made sense to you?"), and 12 ("How often did you feel that you had experiences that challenged you to grow and become a better person?"). Another large international study reported 10 out of 36 countries reaching full scalar invariance (Žemojtel-Piotrowska et al., 2018). These results provide some initial basis to make meaningful comparisons across age and sex. However, researchers should be careful when making crosscultural conclusions.

As noted above, MHC-SF can serve as a diagnostic tool. The results can be categorized into three levels of well-being: flourishing (those who rate 5 or 6 on at least 1 EWB item and at least 6 SWB and PWB items), languishing (those who rate 1 or 2 on at least 1 EWB item and at least six SWB and PWB items), and moderate (those who are neither flourishing nor languishing). Among the 19 samples that reported the categorical diagnosis, 15 roughly demonstrated a similar pattern, with 32.9 % flourishing, 57.0 % moderate, and 8.9 % languishing on average. These results suggest that the MHC-SF may be useful as a diagnostic tool, given that they reveal similar norms across samples. However, two Korean samples (one on adults and the other on high-school students) demonstrated much lower flourishing (8.0 % and 11.7 %, respectively) compared to other samples (Lim et al., 2013; Lim, 2014), whereas a U.S. university student sample and a Canadian adult sample demonstrated much higher percentages of flourishing (51.8 % and 76.9 %, respectively; Gilmour, 2014; Keyes et al., 2012). These differences may suggest that the MHC-SF is useful at revealing potential differences in flourishing, or that there are significant differences in how different cultures use the MHC-SF.

Summary

Overall, the MHC-SF demonstrates good reliability (internal consistency and temporal reliability) and validity evidence for both the overall scale and individual subdimensions. It serves as a useful diagnostic tool to identify individual levels of flourishing. The findings well support a three-factor structure, although more recent studies reported a bet-

ter fit of a bifactor model, suggesting an overarching psychological flourishing dimension informed by the content of the specific dimension. Once this general dimension is accounted for, the specific factors from the subscales do not account for substantial reliable variance beyond that. Future researchers should seek to better understand the role of the MHC-SF total score, as the findings suggested higher reliability compared with the individual subscales. In addition, the subjective well-being dimension consistently shows lower levels of reliability and validity coefficients, especially when adopted to non-Western cultures. Moreover, while the MHC-SF generally shows measurement equivalence on age and sex, there is less evidence of measurement equivalence across countries. Care needs to be taken when using the MHC-SF to compare psychological flourishing across cultures.

PERMA-Profiler

To provide insight into identifying the building blocks of flourishing, Seligman (2011) identified five pillars and proposed the PERMA model in Seligman's Well-Being Theory: Positive emotions (subjective experience of happiness for the past, present, and future), Engagement (the use of the force of character and an individual's talents and capacity), Relationships (creativity and altruism in social relations), Meaning (achieving purpose in life), and Accomplishment (striving for success and victory in the pursuit of self-realization). Empirical research has shown evidence of each component contributing to a variety of life domains, including resilience (Tugade & Fredrickson, 2004), life satisfaction (Kashdan et al., 2009), and reduced risk of depression (Manderscheid et al., 2010).

Following this model, the PERMA-Profiler was developed to measure these five elements, along with negative emotions and health (Butler & Kern, 2016). The measure consists of 23 items in total: 15 items were designed to cover the five core dimensions (3 items each), with 8 filler items on health (3 items), negative emotions (3 items), loneliness (1 item), and overall happiness (1 item). Sample items include "In general, how often do you feel joyful?" and "To what extent do you feel loved?" rated on an 11-point scale ranging from 0 to 10. Overall psychological flourishing is calculated by averaging all the responses, which range from below 5 (*languishing*) to 9 and above (*very high functioning*). Our initial search rendered 14 articles, all of which were included in the current review.

Reliability

The main psychometric properties of the scale came from the development studies by Butler and Kern (2016), who recruited eight English-speaking samples internationally to validate the scale. The mean Cronbach's alpha was .94, with all samples except for one reporting the alpha above .90. All dimensions showed acceptable internal consistencies. The mean Cronbach's alpha was .88 (.84 to .89 range) for Positive emotions, .72 (.60 to .80 range) for Engagement, .82 (.75 to .85 range) for Relationships, .90 (.85 to .91 range) for Meaning, and .79 (.70 to .84 range) for Achievement. Additionally, the mean Cronbach's alpha was .71 (.71 to .77 range) for Negative emotions and .92 (.92 to .94 range) for Health. We also examined three other articles that provided reliability information with the PERMA English version. All studies reported a Cronbach's alpha higher than

.90. All individual PERMA dimensions exhibited good internal consistency, with Cronbach's alphas ranging from .66 to .88.

The test-retest correlations reported in the development studies ranged from .69 (over 1 year) to .88 (over 2 weeks) for the overall PERMA score (Butler & Kern, 2016). The individual dimensions reported acceptable temporal reliability as well. Test-retest correlations ranged from .51 (Engagement over 1 year) to .90 (Relationships over 2 weeks), with 93.3% over .60. In general, these results support that the PERMA-Profilier appears to show a high level of internal consistency and stability over time.

Validity

The development studies confirmed the five-factor proposed structure across all online adult samples from the US, the UK, Australia, Hong Kong, and Malaysia. This provided initial support for the factorial structure of the scale. However, there occurred to be some issues when applying to other cultural contexts, which is covered in the following section.

Regarding the criterion-validity evidence of the measure, the overall PERMA factor correlates positively with flourishing ($r = .84$), health ($r = .50$), and life satisfaction ($r = .78$) in the development studies. These correlations are expected and consistent, which reflect the concept of psychological flourishing that emphasizes positive experiences. The measure was also negatively correlated to psychological symptoms including anxiety ($r = -.50$), perceived stress ($r = -.55$), and depression ($r = -.61$). Ryan et al. (2019) reported that, while the construct correlated moderately to strongly with subjective measures ($r = .63$ with mental health, $r = -.65$ with depression, $r = -.37$ with anxiety, and $r = -.46$ with stress), it correlates only negligibly ($r = -.03$ with objective activity and $r = -.05$ with sleep) or not at all significantly with objective ones ($r = .15$ with physical health).

Individual dimensions also demonstrated acceptable convergent validity evidence (Butler & Kern, 2016; Ryan et al., 2019). Positive emotions were found to be positively correlated with self-acceptance ($r = -.75$ to $-.73$) and negatively correlated with depression ($r = -.75$ to $-.49$), anxiety ($r = -.53$ to $-.27$), perceived stress ($r = -.58$ to $-.28$), and negative emotions ($r = -.61$ to $-.29$). Engagement reported weak but statistically significant correlations with compassion ($r = .25$), activist identification ($r = .18$), and work performance ($r = .25$). Relationships were found to be positively correlated with social support ($r = .50$ to $.68$) and negatively correlated with loneliness ($r = -.63$ to $-.50$). Meaning was positively correlated with purpose in life ($r = .30$) and self-acceptance ($r = .45$), and Achievement with self-efficacy ($r = .65$) and less burn-out ($r = .57$).

Application in Diverse Samples

The original English version has been validated in different cultural contexts, including in Australia, the United States, the United Kingdom, Hong Kong, Canada, etc. It has been translated into more than 10 languages, including German, Turkish, Italian, and Korean (see <https://www.peggykern.org/questionnaires.html> for many of these scales). Five articles provided reliability information from international samples, including Indonesia, Italy, Greece, etc. (e.g., Hidayat et al., 2018; Giangrasso, 2018; Pezirkianidis et al., 2021),

all of which reported a Cronbach's alpha of .80 and above. In these articles, we note that Engagement (.56 to .84) showed slightly lower reliabilities than the other dimensions (.77 to .90 for Positive Emotions, .70 to .96 for Relationships, .78 to .91 for Meaning, and .70 to .93 for Achievement). Test-retest reliabilities reported were .81 over 2 weeks (Ayse, 2018) and .88 over 1 month (Watanabe et al., 2018).

We examined the convergent and discriminant validity evidence, which was supported in three international samples covering German-speaking adults (Wammerl et al., 2019), Italian university students (Giangrasso, 2018), and Greek adults (Pezirkianidis et al., 2021). The overall score was found to correlate positively with subjective happiness in an Italian university student sample ($r = .81$) and psychological well-being in the Greek adult sample ($r = .77$). It negatively correlated with depression in the German-speaking adult sample ($r = -.76$). External validity evidence as the correlation coefficients reached $-.59$ between Positive Emotions and negative affect and $.61$ between Achievement and environmental mastery (Wammerl et al., 2019). Giangrasso (2018) also reported positive correlations between Relationships and positive relationships ($r = .68$) and between Meaning and purpose in life ($r = .82$). Nonetheless, Pezirkianidis et al. (2021) used confirmatory factor analysis to examine the discriminant validity evidence of the different PERMA dimensions and found that the Engagement factor was not well distinguished from other dimensions. This appears to be aligned with the findings that Engagement also has lower reliability than the other dimensions in international samples.

Applied to a diverse cultural context including India, Greece, Germany, Austria, Italy, and Turkey, most of the eight factorial validation articles supported the intercorrelated five-factor structure of the originally proposed model (e.g., Ayse, 2018; Giangrasso, 2018; Pezirkianidis et al., 2021). Nevertheless, there were instances where the results failed to extract the proposed five factors. Two articles, with physically inactive Australian adults and U.S. student veterans respectively, suggested a two-factor solution in further exploratory factor analyses (Ryan et al., 2019; Umucu, Grenawalt et al., 2019). One article, with English-speaking Malaysian adults, revealed a three-factor solution according to the Kaiser-Guttman criterion (Khaw & Kern, 2015). We note that researchers also hypothesized and explored other factor structures (such as higher-order structure and bifactor structure) (e.g., Hidayat et al., 2018; Wammerl et al., 2019). In general, there appears to be support for the structure of the PERMA-Profil in a variety of cultural contexts, though specific studies may not show the five-factor structure.

We found two articles that tested measurement invariance of the scale, both supporting full scalar invariance of sex (Pezirkianidis et al., 2021; Wammerl et al., 2019). The former article also reported full scalar invariance across age groups. The latter tested the German version across three countries. A strict invariance model was reached between Austria and Germany (the Swiss data were excluded from analysis because of the small sample size). These results provided preliminary evidence that the PERMA-Profil might be interpreted similarly across sexes and age groups. However, more research is necessary to comprehensively understand the measurement equivalence across cultures, as the factor model only converges for some samples.

Summary

Across multiple studies, there is evidence of the reliability and validity of the PERMA-Profiler. However, some researchers also pointed out the potential issues (e.g., lower reliabilities, low discriminant validity evidence with other dimensions) with the Engagement dimension when applied to diverse cultural contexts. Future researchers are encouraged to test the measurement equivalence of the measure and to validate its factor structure in international contexts as there is competing evidence that failed to replicate the five-factor model. In terms of its usage, Wammerl et al. (2019) noted that the PERMA theory is insightful for addressing at least some of the building blocks of subjective well-being, which might be useful for developing interventions in the context of psychotherapy, coaching, or counseling.

Comprehensive Inventory of Thriving and Brief Inventory of Thriving

The Comprehensive Inventory of Thriving (CIT) was developed rather recently to “measure a broad range of psychological well-being constructs and represent a holistic view of positive functioning” (Su et al., 2014). The concept of “thriving” was used to connote a comprehensive image of well-being beyond the traditional split between the hedonic and eudaimonic framework. The CIT integrated key hedonic and eudaimonic models, including subjective well-being (Diener, 1994), psychological well-being (Ryff & Keyes, 1995), self-determination theory (Ryan & Deci, 2000), and the PERMA theory of well-being (Seligman, 2011). Based on these previous well-being frameworks, researchers identified 18 subscales under seven dimensions as the overarching theoretical framework: Relationships (Support, Community, Trust, Respect, Loneliness, and Belongingness subscales), Engagement (Engagement subscale), Mastery (Skills, Learning, Self-Efficacy, Self-Worth, and Accomplishment subscales), Autonomy (Lack of Control subscale), Optimism (Optimism subscale), Subjective well-being (Life Satisfaction, Positive Emotions, and Negative Emotions subscales), and Meaning (Meaning subscale). The CIT was built upon this structure with 3 items under each subscale, for a total of 54 items. Respondents rate their agreement on a 5-point scale from 1 (*strongly disagree*) to 5 (*strongly agree*). Su et al. (2014) presented scale norms in a large adult sample within the U.S. We identified nine recent studies validating the scale (and/or its variant) and presented the key psychometric properties below.

Reliability

In the development studies, the Cronbach’s alphas ranged from .71 to .96 across samples (one college student sample, one elder adult sample, one lower-income adult sample, and two adult samples) and dimensions (Su et al., 2014). Test-retest reliabilities were around .60 and above over 4 months for all subscales. In general, the CIT scale showed good internal consistency and test-retest reliability.

Validity

The 18-factor correlated model had an excellent model fit and was thus supported in the original studies (Su et al., 2014). This indicated the correlated, yet distinct, dimensions assessed by the subscales. As for convergent validity evidence, CIT was found to correlate moderately to strongly with extant measures: flourishing ($r=.30$ to $.73$ range), self-mastery ($r=.26$ to $.69$ range, the highest with the Self-Efficacy subscale), Optimism ($r=.26$ to $.82$ range, the highest with the Optimism subscale), satisfaction with life ($r=.20$ to $.90$, the highest with the Life Satisfaction subscale), and core self-evaluations ($r=.26$ to $.78$ range). In line with theory, it negatively correlated with scales that measure psychological symptoms, including PHQ-9 ($r=-.59$ to $-.18$ range) and GAD-7 ($r=-.51$ to $-.12$ range). Su et al. (2014) also examined the incremental validity of the CIT compared with Satisfaction with Life Scale (SWLS; Diener, Emmons, Larsen, & Griffin, 1985), Flourishing Scale (Diener et al., 2010), Life Orientation Test-Revised (Scheier, Carver, & Bridges, 1994), Self-Mastery Scale (Pearlin & Schooler, 1978), and Core Self-Evaluations Scale (Judge, Erez, Bono, & Thoresen, 2003). Results showed an average of 59.63% in the additional variance accounted for across health-related outcomes (existence of medical conditions, level of physical functioning, etc.), suggesting its utility beyond the existing measures.

Application in Diverse Samples

Eight translations of the scale can be found on Dr. Ed Diener's personal website (http://labs.psychology.illinois.edu/~ediener/CIT_BIT.html). Research has validated the scale in a diverse cultural background, including German, Italian, Chinese, Turkish, and Spanish contexts covering 14 countries. Regarding reliability, in an international study across several countries, including the United States, Argentina, Australia, China, Germany, India, Spain, Singapore, Turkey, Mexico, and Russia, Cronbach's alphas were above .70 for all dimensions except for Engagement (.37 in Argentina, .37 in Mexico and .54 in Spain; Wiese et al., 2018). In another Brazilian sample, the Cronbach's alphas ranged from .70 to .95 (Martins & Ferreira, 2018). We did not find any article reporting the test-retest reliability of the CIT in other cultures.

Convergent validity evidence of both the general scale and the subscales was established in the Chinese and Brazilian samples. For example, Duan et al. (2020) reported positive correlations between the CIT overall score with life satisfaction ($r=.57$) and flourishing ($r=.68$); negative correlations were found with depression ($r=-.46$), anxiety ($r=-.34$), and stress ($r=-.35$). As for the subscales, strong correlations were reported in expected directions, Life Satisfaction with life satisfaction ($r=.78$), Optimism with optimism ($r=.57$), Positive feelings with flourishing ($r=.73$), etc. CIT was also positively correlated to students' life satisfaction in a sample of Italian elementary-school students ($r=.16$ to $.56$ across subscales; Andolfi et al., 2017).

The correlated 18-factor model showed satisfactory model fit indices (compared with single-factor, bifactor, seven-factor, or 18 first-order factors with seven second-order factors models) in a Chinese community sample (Duan et al., 2020), a Brazilian adult sample (Martins & Ferreira, 2018), and international samples in five out of eight countries (Wiese et al., 2018). The three countries that failed to replicate the factor structure were

Argentina, China, and Mexico, with no acceptable alternative solution. Strict measurement invariance was reported across eight out of 11 countries, further suggesting the measurement equivalence across cultures (Wiese et al., 2018). However, higher-order structure models attained the best fit in a study with German-speaking adults (Hausler et al., 2017) and a study with Italian children (Andolfi et al., 2017). In sum, though some minor disparities in factor models do occur, the multidimensionality of the scale is supported across studies. Further, the CIT measure appears to show measurement equivalence across nations.

BIT: Variant of CIT

Ten core items were selected from the CIT to develop the Brief Inventory of Thriving (BIT), to serve as a comparatively shorter screening tool for mental health status and as an index of general psychological flourishing. We examined five articles that tested its psychometric properties. Cronbach's alphas ranged from .75 to .93 across samples, suggesting good internal consistency of the scale. The one-factor structure was confirmed in all studies, while results suggested error correlations for Turkish and Russian samples (Wiese et al., 2018).

Two articles tested measurement equivalence across countries of the BIT. Sorgente et al. (2018) reported full configural, full metric, partial scalar, and partial unique invariance across Italian, Portuguese, and Chinese young adults. Further, Wiese et al. (2018) reported full configural, full metric, and partial scalar invariance across 11 countries, including Argentina, Australia, China, Germany, India, Mexico, Russia, Singapore, Spain, Turkey, and the United States. These results again support the hypothetical monodimensional structure of the BIT and expand its applicability over diverse cultural backgrounds.

For convergent validity evidence, BIT was found to positively correlate with flourishing, life satisfaction, optimism, self-mastery, meaning in life, and core self-evaluations. Negative correlations were found between BIT and negative emotions, depression, anxiety, and stress. These results provided discriminant and convergent validity evidence of the scale. In terms of incremental validity, it explained an average of 29.48% additional variance over other established scales in most health outcomes. More importantly, it improved upon FS in predicting the health outcomes by 20.08%, suggesting its unique predictive ability is different from FS (Su et al., 2014). In a Chinese community sample, BIT was found to be the only significant contributing factor (together with FS and SWLS) in explaining the variance of ill-being (depression, stress, and anxiety) in the risk group. These results suggest BIT's strong predictability of various behavior and health outcomes, enlarging its breadth of application.

Summary

There appears to be good reliability and validity evidence for the CIT and especially for its shorter variant, the BIT. There is also evidence of its incremental validity beyond extant measures for predicting outcomes such as health. However, the Engagement dimension showed lower reliabilities when applied to other cultures (Wiese et al., 2018). Interestingly, we observed similarly lower reliabilities of the Engagement subscale in the

PERMA-Profiler. Future efforts are encouraged to delve deeper into this aspect. Further, there have been relatively large-scale studies to examine the use of the CIT and BIT across multiple different countries, generally showing measurement invariance (except for Argentina, China, and Mexico).

Flourishing Scale (FS)

The Flourishing Scale (Diener et al., 2010) is a brief measure designed to capture optimal human flourishing across reasonably comprehensive domains. Rather than separating into individual facets, the FS consists of eight items providing an overall score of social-psychological functioning, including meaning and purpose in life, relationships, engagement, competence, and optimism. Although the FS was developed to be unidimensional in terms of its factor structure, it encompassed content from multiple sources and conceptualized psychological flourishing as a construct contributed by various psychosocial dimensions. This seems analogous to the MHC-SF bifactor models that seek to extract a general factor from the different subscales. Considering its integrative nature and economy of time in its application, we decided to include the scale in the current review to provide future researchers and practitioners with a useful measurement tool to consider when seeking to broadly assess psychological flourishing.

Sample items of the FS include “People respect me” and “I am a good person and live a good life,” rated on a 7-point scale ranging from 1 (*strongly disagree*) to 7 (*strongly agree*). An overall flourishing score is calculated by summing up all the responses, with higher scores implying higher levels of psychological flourishing. We found 30 articles and included 26 of them in our final review.

Reliability

The alpha reliability was .87, and the test-retest correlation over 1 month was .71 in the original paper on U.S. college students (Diener et al., 2010). Howell and Buro (2015) reported a similar alpha of .89 in Canadian college students. The following studies covering diverse samples reported similar results, which are described in the following section. In sum, the statistics provided initial support for the satisfactory internal consistency and temporal stability over 1 month of the FS.

Validity

The FS generally displayed good validity evidence in Western cultures. It showed positive correlations with Cantril’s Ladder ($r=.57$; Diener et al., 2010) and subjective happiness ($r=.67$; Hone et al., 2013). Positive correlations with life satisfaction were also reported ($r=.62$ and $.64$, respectively). The FS was negatively correlated with loneliness ($r=-.28$; Diener et al., 2010), anxiety ($r=-.65$), depression ($r=-.65$), and stress ($r=-.60$; Umucu, Grenawalt et al., 2019). In terms of factorial validity evidence, results showed a consistent one-factor structure across the five articles on Western populations, which agreed upon its overarching theoretical framework. However, some minor issues oc-

curred, which are described in more detail below. In all, the results showed satisfactory psychometric properties as a measure of flourishing.

Application in Diverse Samples

The FS has been translated into 25 languages to date (<https://eddiener.com/scales/9>). The original English version was developed and validated in the United States and Singapore (Diener et al., 2010). Following studies validated this scale in countries over the globe targeting different populations, including Japanese college students (Sumi, 2014), Indian adolescents (Singh et al., 2016), and patients with chronic back pain (Perera et al., 2018). All 30 studies demonstrated satisfactory reliability, reporting a Cronbach's alpha equal to or greater than .80 (range .80 to .95), except for one study of Spanish parents with children of cancer reporting an alpha of .74 (Pozo Muñoz & Bretones Nieto, 2019) and another with Portuguese university students of .78 (Silva & Caetano, 2013). These results together supported the reliability of the FS across international contexts.

The FS scores correlated with measures of well-being in diverse samples among Italian (Giuntoli et al., 2017), Russian (Didino et al., 2019), Egyptian (Salama-Younes, 2017), and Portuguese adults (Silva & Caetano, 2013). Among college students and adults in Brazil, FS correlated positively with positivity at $r = .65$ level (Fonseca et al., 2015). Further, among Greek adults (Kyriazos et al., 2018) FS was positively correlated with gratitude ($r = .47$) and resilience ($r = .35$). In a sample of Chinese adults, the FS showed positive correlations with virtues ($r = .55$), including relationship ($r = .44$), vitality ($r = .38$), and conscientiousness ($r = .49$; Tang et al., 2016). Therefore, there is good validity evidence for the FS scale across different countries.

Across demographic groups, configural invariance was supported across ages (adolescents vs. adults), majors, employment status (employed vs. unemployed), and administration methods (online vs. paper; Giuntoli et al., 2017; Singh et al., 2016; Villieux et al., 2016). Scalar equivalence was also reported in a Spanish sample between two universities (De la Fuente et al., 2017) and in a Greek adult sample across sexes (Kyriazos et al., 2018).

Although the FS has been applied in many countries, we did not find studies conducted to examine the measurement equivalence of FS across nations or cultures. Some initial evidence of measurement equivalence comes from the FS factor structure across diverse samples. There seems to be strong support for the one-factor solution across samples, suggesting at least configural equivalence. However, numerous studies required the specification of error covariances between items to reach satisfactory model fit, suggesting that there may not be scalar or metric invariance (e.g., Didino et al., 2019; Hone et al., 2013; Kyriazos et al., 2018; Perera et al., 2018; Tong & Wang, 2017; Umucu, Grenawalt et al., 2019). We suggest that more research on cultural equivalence is necessary before making a conclusion.

Summary

In general, the FS showed good reliability and validity evidence. It has been translated and used in many countries and shows a robust one-factor solution to capture psychological flourishing. From the evidence presented so far, it appears that the FS is a useful tool to integratively measure psychological flourishing, especially considering its length and breadth. Future researchers should seek to examine whether the FS is measurement invariant across nations to determine whether FS scores are comparable across cultures and languages.

Psychological Well-Being Scale

At a time when most researchers understood well-being from a hedonic perspective, Ryff argued for the importance of psychological well-being, which should include positive human functioning beyond just experiencing positive feelings (Ryff, 1989). A scale was developed to measure psychological well-being from six aspects, including Autonomy (AU), Environmental Mastery (EM), Personal Growth (PG), Positive Relations with others (PR), Purpose in Life (PL), and Self-Acceptance (SA; Ryff, 1989). The initial inventory had 120 items in total. The scale established various shorter versions afterward with modifications and development. The most used versions have a length of 84 items (14 items per subscale), 54 items (9 items per subscale), or 18 items (3 items per subscale). To adapt it to diverse cultural and linguistic contexts, researchers also developed various other forms of different item selection based on the original inventory. Generally, the scale is rated on a 6-point Likert-type scale from 1 (*strongly disagree*) to 6 (*strongly agree*). In our current review, we summarized results from a total of 45 studies that validated the scale.

Reliability

The original PWBS inventory (120 items) was tested on 321 adults and showed good internal consistency across subscales (Ryff, 1989). Cronbach's alphas ranged from .86 to .93 across subscales. The test-retest correlations ranged from .81 to .88 over a 6-week interval, establishing acceptable temporal consistency. A shortened form (84 items) was then developed and tested on midlife and aging adults. Cronbach's alphas reported similar results, ranging from .82 to .91 across subscales (Ryff & Essex, 1992; Ryff et al., 1994; Schmutte & Ryff, 1997). However, the subscales showed a trend of decreasing reliability with a smaller number of items. While the 18-item version showed acceptable reliability in the overall score, reporting an alpha of .81 (Keyes et al., 2002), the subscales only reported low alphas ranging from .33 to .59 (Keyes et al., 2002; Ryff & Keyes, 1995). While the longer forms of PWBS reported acceptable internal consistency, the results raised concern over the ability of items to reliably measure the subscales in the 18-item version.

Validity

Four studies reported validity evidence for the PWBS (Keyes et al., 2002; Ryff, 1989; Ryff et al., 1994; Ryff & Keyes, 1995). Most subscales showed moderate to strong correlations with life satisfaction ($r = .35$ to $.73$), except for PG ($r = .18$ to $.38$) and AU ($r = .12$ to $.30$). In general, the subscales displayed lower correlations with hedonic well-being measures, including positive affect ($r = .19$ to $.50$) and subjective happiness ($r = .08$ to $.54$), which aligned with the hypothesis. Individual subscales also showed acceptable convergent validity evidence with relevant external measures. For example, SA positively correlated with self-esteem at $r = .62$, EM with internal control at $r = .52$, and PL with morale at $r = .55$ (Ryff, 1989). PR, AU, and PG again showed relatively lower correlations with other indexes. Ryff (1989) pointed out that these three dimensions were not well represented at the time of research. Indeed, the following research provided more convergent validity evidence, which is covered in the following section.

Regarding factorial validity evidence, the proposed factor structure by Ryff (1989) was a correlated six-factor model. As mentioned earlier, Ryff and Keyes (1995) also introduced a hierarchical second-order factor as all the subscales theoretically contribute to general psychological well-being. Only one study tested the factor structure on the 120-item version of PWBS and failed to replicate the *a priori* 6-factor model (Kafka & Kozma, 2002). Further research also revealed competing results in 42- and 54-item versions. For example, Boers (2014) reported an acceptable 6-factor structure, while other studies reported the opposite (Abbott et al., 2006; Burns & Machin, 2009). On the contrary, the 18-item version best supported the proposed factor models, with all five articles confirming the existence of six first-order factors corresponding to the *a priori* dimensions. These results raised questions over the factorial structure of the scales, especially the longer forms. We compare more results from diverse samples in the following section as well.

Application in Diverse Samples

Ryff's PWBS has been translated into multiple languages and applied in more than 20 countries over the globe. The trend persisted so that the subscales showed lower reliability with smaller numbers of items. Specifically, the 18-item version reported Cronbach's alphas ranging from $.17$ to $.68$ (90.0% under $.60$) in four studies. An exception was with a sample of Swedish white-collar workers (median = $.65$), where the only subscales with low reliability were PL ($.24$) and AU ($.53$; Lindfors et al., 2006). Among all, PL showed the lowest reliability ($.17$ to $.33$) compared to other dimensions.

Nevertheless, the longer forms of the PWBS showed comparatively higher reliabilities. Of the nine articles that provided reliability information for the 54-item version, most reported moderate to high Cronbach's alpha ($.39$ to $.84$, 86.3% were over $.60$), except for Hong Kong adults ($.39$ to $.51$; Cheng & Chan, 2005) and Iranian university students ($.53$ to $.68$; Shokri et al., 2008).

The 84-item version also demonstrated moderate to acceptable reliability across 16 studies ($.37$ to $.96$), with 83.3% of alpha statistics over $.60$. More specifically, the results showed relatively lower reliability in Japanese university students ($.45$ to $.83$; Kitamura et al., 2004), Hong Kong university students ($.55$ to $.70$; Cheng & Chan, 2005), Iranian university students ($.57$ to $.76$; Bayani et al., 2008), and Portuguese adolescents ($.36$ to

.50; Fernandes et al., 2010). Test-retest reliability across a 1-month interval was established with Turkish university students (.78 to .97; Akin, 2008) and Italian university students (.78 to .82, except for .21 for Autonomy and .31 for Environmental Mastery; Ruini et al., 2003); 2-month temporal reliability (.70 to .82) was established with a sample of Iranian undergraduates (Bayani et al., 2008).

Across different samples around the world, the overall score and the individual dimensions of the scales were found to correlate moderately to strongly with life satisfaction, positive affect, and happiness (e.g., Chan et al., 2019; Akin, 2008). A negative relationship was found with negative affect and depression (Machado et al., 2013). More specifically, in a Hong Kong adolescent sample, SA was found to correlate strongly with self-esteem (.61), EM with self-efficacy (.56), PG with personal growth initiative (.54), PL with the presence of meaning in life (.65), AU with adolescent autonomy (.57), and PR with social self-efficacy (.51; Chan et al., 2019). These findings provide convergent validity evidence of both the general scale and the individual dimensions.

The issues with the factor structure of the longer-form PWBS (i.e., 84-, 54-, and 42-item versions) persisted in non-Western cultures. The six-factor structure was supported in nine articles (e.g., Costea-Bărluțiu et al., 2018; Freire et al., 2019; Shokri et al., 2008), while another 12 articles failed to report acceptable model fit indices (e.g., Burns & Machin, 2009; Van Dierendonck, 2004; Villarosa & Ganotice, 2018). The 18-item version again showed better results. Eight out of 10 studies found acceptable model fit indices for the proposed six-factor structure.

A potential issue that researchers commonly identified is the overlap among PG, PL, EM, and SA in the longer forms (van Dierendonck, 2004; Springer & Hauser, 2006). These four dimensions typically showed moderate to high intercorrelations, with approximately 66.7% over .60 out of 22 studies. These results suggest that items from these four dimensions might contribute to an underlying construct. Moreover, factor analyses indicated that items from these dimensions had significant crossloadings on other factors (Abbott et al., 2006; Burns & Machin, 2009; Sirigatti et al., 2009; Tomás et al., 2008; van Dierendonck et al., 2008).

Three studies tested measurement invariance on gender groups. Two supported configural, metric, and scalar invariance across sexes, with the other one on UK adults suggesting that metric invariance did not hold, as men and women exhibited different factor loadings (Guidon et al., 2005). One study reported full metric invariance between Italy and Belarus (Sirigatti et al., 2013). Another study suggested a similar factor structure between Australian and international teacher samples, but the Norwegian sample differed from the previous two (Burns & Machin, 2009). Because research testing the relevant field is relatively scarce, future research should devote more to establishing the equivalence across samples.

Summary

The Ryff's PWBS generally showed satisfactory psychometric properties. From the research examined, there seems to be a trade-off with the length of the scale. The 18-item version showed better factorial validity evidence than the longer forms. However, with fewer items, there appear to be more challenges with reliability, particularly on the PL